Studying soil moisture at a national level through statistical analysis of NASA NLDAS data

Gonzalo E. Espinoza-Dávalos, David K. Arctur, William Teng,
David R. Maidment, Irene García-Martí and Georges Comair

ABSTRACT

The purpose of this research is to enable better understanding of current environmental conditions through the relations of environmental variables to the historical record. Our approach is to organize and visualize land surface model (LSM) outputs and statistics in a web application, using the latest technologies in geographic information systems (GISs), web services, and cloud computing. The North American Land Data Assimilation System (NLDAS-2) (http://ldas.gsfc.nasa.gov/nldas/; Documentation: ftp://hydro1.sci.gsfc.nasa.gov/data/s4pa/NLDAS/README.NLDAS2.pdf) drives four LSM (e.g., Noah) (http://ldas.gsfc.nasa.gov/nldas/NLDAS2model.php) that simulate a suite of states and fluxes for central North America. The NLDAS-2 model output is accessible via multiple methods, designed to handle the outputs as time-step arrays. To facilitate data access as time series, selected NLDAS-Noah variables have been replicated by NASA as point-location files. These time series files or 'data rods' are accessible through web services. In this research, 35-year historical daily cumulative distribution functions (CDFs) are constructed using the data rods for the top-meter soil moisture variable. The statistical data are stored in and served from the cloud. The latest values in the Noah model are compared with the CDFs and displayed in a web application. Two case studies illustrate the utility of this approach: the 2011 Texas drought, and the 31 October 2013 flash flood in Austin, Texas.

Key words | cloud processing, data rods, NLDAS, soil moisture, web services, web applications

Gonzalo E. Espinoza-Dávalos
(corresponding author)
David K. Arctur
David R. Maidment
University of Texas at Austin,
Austin,
TX 78712,
USA
E-mail: gespinoza@utexas.edu

William Teng
NASA Goddard Earth Sciences Data and
Information Services Center,
Greenbelt,
Maryland,
USA

Irene García-Martí
University of Twente,
Drienerlolaan 5,
Enschede 7522 NB,
The Netherlands

Georges Comair
Suez Environnement,
France

ABBREVIATIONS

API	application programming interface
CDF	(statistical) cumulative distribution function
CRWR	University of Texas Center for Research in Water Resources
DEM	digital elevation model
GCM	global circulation model
GIS	geographic information system
GLDAS	global land data assimilation system
HPC	high-performance computing
JSON	JavaScript Object Notation
LDAS	Land Data Assimilation System
LSM	land-surface model
NASA	National Aeronautical and Space Administration
NLDAS	North American Land Data Assimilation System

Noah	acronym formed by the four institutions that developed it: National Centers for Environmental Prediction, Oregon State University, Air Force, and Hydrology Laboratory of the National Weather Service
REST	REpresentational State Transfer
SDK	software development kit
TACC	(University of) Texas Advanced Computing Center
VIC	Variable Infiltration Capacity model

INTRODUCTION

Hydrologic variables are used to describe spatial-temporal processes that are the result of intricate interactions from

large scale climate dynamics to local conditions. This complex system carries a large degree of uncertainty, so hydrologic variables are considered stochastic, i.e., variables with underlying distributions that vary in time and space. The Land Data Assimilation System (LDAS) is a model developed, implemented, and shared by the National Aeronautics and Space Administration (NASA) with the purpose of being a thorough, consistent dataset of the heat and water fluxes between the atmosphere and the land-surface hydrology. LDAS variables are exposed through two sets of products: National (NLDAS) and Global (GLDAS). The NLDAS grid has 0.125 degree spatial resolution (approximately 14 km), with hourly and monthly estimates since 1979. The GLDAS grid has 0.25 degree resolution (approximately 28 km), with 3-hourly and monthly estimates since 2000. The LDAS model output carries information about the spatial-temporal nature of the hydrologic variables, and provides a consistent and comprehensive time-space dataset for the model variables represented. Through statistical analysis of the 35-year record of these simulations, it is possible to gain important insights into seasonal patterns, spatial distribution, and frequency values of the model output.

Gourbesville (2009) describes a general context of how new technologies can create datasets with added value, emphasizing that current datasets are more extensive and dense than the ones in the past. It concludes that future studies must focus on data management; improving the ways that information is shared and presented. The latest conditions of soil moisture in NLDAS can be compared to those in the past by using statistical distributions for each variable unique to each calendar day and spatial grid point.

Lakshmi (2004) identifies the challenges in hydrologic sciences for the estimation of variables in ungauged basins. The research acknowledge the importance of satellite products and land-surface models for improving estimations, and emphasizes the need of studying and mapping soil moisture for accurate water balances (Kothyari *et al.* 2010).

Kothyari *et al.* (2010) develop a geospatial application for the analysis of soil erosion rates. The research uses as a study case the Garhwal region in the Himalayas. The study reinforces the fact that geospatial applications efficiently communicate information through maps displaying the geographic distribution of the variable of interest (soil erosion in that case). The research did not provide a web-based application solution, whose implementation would be preferable. In addition, there is a significant effort being applied in the data preparation and post-processing. In the present research, an integration of historic data, statistical analysis, and latest results helps to create a more complete picture of the current state of a hydrologic variable in a spatial location. This integration can be automated, through web applications and the latest developments in geoinformatics technologies to increase the added value of hydrologic data. The objective of the present research is to integrate the latest and historic NLDAS soil moisture data through a statistical analysis, and to expose the results in a web application. This approach has been extended to other NLDAS Noah variables (precipitation rate, snowfall rate, surface runoff, evapotranspiration, and surface temperature) but the visualization and analysis of these variables' time series model outputs is still in progress and is not presented here.

The North American Land Data Assimilation System

The North American Land Data Assimilation System (NLDAS) is a compilation of land-surface model (LSM) datasets of hydrologic variables in a time-space continuum with regular intervals (Mitchell 2004; Xia *et al.* 2012). The strength of NLDAS relies on the forcing parameters (derived from observations) to reduce the bias and error of geophysical models, rather than being coupled with atmospheric models. The advantages of LSM in hydrology over global circulation models (GCMs) are described by Jiang *et al.* (2013). They show that GCMs have problems replicating extreme precipitation values, especially high precipitation values in short periods of time. The research states that the use of these models is improper for flood and drought assessments and suggests that LSM, uncoupled from atmospheric models, and calibrated with *in-situ* measurements, has improved results.

From the models in NLDAS, the Noah model (acronym formed by the four institutions that developed it: National Centers for Environmental Prediction, Oregon State University, Air Force, and Hydrology Laboratory of the National Weather Service) is used in this research. The Noah LSM was derived from a less complex model developed at

Oregon State University in the 1980s which has been subject to revisions and improvements since then (Ek *et al.* 2003). The NLDAS-Noah dataset is indexed by space in the Grid Application Development Software (GrADS) (Mitchell 2004), which means that a file with the spatial coverage over the whole domain can be obtained for a single time-step (Berman *et al.* 2001). Similarly, the NLDAS-Noah data are also available in a web service indexed by time, called 'data rods' (Rui *et al.* 2013), which means that the time-series for a single point in space can be obtained. The double indexing by time and space optimizes the process of accessing the data. Jones *et al.* (2000) pose a methodology to handle temporal data in hydrologic models or applications. This solution improves the performance of the web application prototype in the present research; specifically, at the data access process using the 'data rods' and at displaying the data in the plots.

The research performed by Elshorbagy & El-Baroudy (2009) acknowledges the importance of soil moisture data as key elements for understanding the hydrologic cycle. It uses data-driven techniques to analyze the soil response to external forcing. The research concludes that models of soil moisture are necessary besides *in-situ* measurements or satellite data, due to the intricate physical process. NLDAS provides extensive validated hydrologic datasets with continuous historical information which can be used as a deterministic model, as well as to identify trends, statistical distributions, and the strength of the relationships between variables. The NLDAS dataset is extensive but it can be queried and accessed online with standard filters, such as location, time, and variable of interest, in an automated way. The study performed by (Lakshmi *et al.* 2004) used soil moisture data from the Variable Infiltration Capacity (VIC) model as an indicator of hydrologic extremes (i.e. floods and droughts) in the upper Mississippi basin. The research shows that the analysis of long-term soil moisture and its anomalies can be used as a precise drought indicator. They identified mean and common ranges of soil moisture values during normal, flood, or drought conditions but did not associate probabilities for these values.

Statistical analysis

The expectation of extreme events is one of the major applications of statistics in hydrology. The research of El Adlouni *et al.* (2008) examines the common distributions used in hydrology and classifies them accordingly to their tail behavior. The research reinforces the importance of the tail behavior for the estimation of extreme events. The study performed by Coles *et al.* (2003) uses rainfall data from coastal and central Venezuela and proves that extreme rainfall events although exceptional can be properly estimated. Hence, constructing the rainfall probability distributions can be relevant in the estimation of these extreme events.

The probability distribution of hydrologic variables can be estimated on the complete time series or on partial, minimum, and maximum series. Beguería (2005) focuses on modeling extreme rainfall events using the series of annual maximum and partial duration. This method requires defining threshold values which can be complicated and subjective to estimate. The percentiles of the distributions are affected by this threshold (Katz *et al.* 2002), that can lead to subjective results. If sufficient data are available, the complete time series can be used to estimate the probability distribution. The research performed by Mishra (2009) recognizes the importance of estimating the range of likely values and uncertainty. In the present research, the uncertainty is measured from the long-term historical probability distributions, the latest values are compared to these distributions to evaluate if the values that are extreme are expected.

The study performed by Husak *et al.* (2007) set the basis of computing cumulative distribution functions (CDFs) per grid cell and per time interval (monthly). It uses rainfall data from the collaborative historical African rainfall model (CHARM) (Funk *et al.* 2003). The CHARM data are coarser than NLDAS with half-degree cell size, and the research downscaled the data using an underlying digital elevation model, which could induce additional errors.

Information technologies and data access

The paper presented by Beniston *et al.* (2012) is a thorough description of the data access challenges faced in water resources research. In summary, the challenges are: (1) sparse datasets; (2) gaps in data; (3) consistent, reliable availability; (4) charges involved on sharing datasets; (5) disparity between data availability in successful socio-economical areas vs. areas with limited resources; and (6) the

inconsistent use of standards in storing and sharing information. The use of NLDAS and the data rods web services minimizes all the challenges because: (1) the dataset is continuous in time and space; (2) there are no gaps of data in the model; (3) the data are available through web services; (4) the dataset is public; and (5) the cell size and time interval are the same for all estimations in the dataset. The remaining challenges (6) are to connect the data services with geographic information and ease the process of exposing the results in informative websites. The present project provides solutions for these challenges, including a web application prototype that uses international data exchange standards and cloud storage.

Choi *et al.* (2005) implements a web application for water management through a spatial decision support system. The research states the advantages of web-based geographic applications to ease the constraints of using large datasets, data transmission, and integration between models. The web application was implemented using a Common Gateway Interface method, instead of a more robust web framework (e.g., Django) as in the present research. The use of web application frameworks can facilitate the web application development process and the interconnectivity with data services.

The emerging field of CyberGIS (Liu *et al.* 2015) is about the evolution of geographic information systems (GISs) on the web. It studies all parts required for a spatial analysis on the web: data services, spatial services, geoprocessing services, online modeling and analysis, and infrastructure. CyberGIS focuses on interactive solutions that rely on large geospatial datasets and its integration with other networks. The web application presented in this research emerges as a significant prototype of CyberGIS for statistical and hydrologic fields.

To summarize, the methodology for this research builds on the latest technologies in high-performance computing (HPC), cloud storage and deployment, and GISs. The analysis is performed on a large dataset (NLDAS), using the soil moisture data of Texas and studying the soil moisture conditions during the 2011 drought and the 'Halloween Flood' on 31 October 2013 in Onion Creek located south of Austin. The historic analysis of NLDAS soil moisture provides relevant information about extreme values and their probability distributions, providing a thorough study of the model output. This approach can indicate which areas may be most vulnerable to hydrologic extreme events, visualizing key aspects of large scientific datasets in informative web applications.

STATISTICAL ANALYSIS

The statistical analysis is based on the modeling of the empirical CDFs on a daily basis, based on the NLDAS 2-Noah soil moisture data obtained for Texas during a period of 35 years (1979–2013). The data retrieval is made through NASA's data rods web service, which provides the time series for a given NLDAS grid cell. The use of data rods improves the data access process for this analysis, because each cell can be processed independently (i.e., in parallel). The data retrieval process is automated and implemented using HPC resources at the Texas Advanced Computer Center (TACC).

First, a subset of soil moisture (variable x) values $S_{x,d}\big|_{t_1}^{t_2}$ is obtained from the time series $S_x\big|_{t_1}^{t_2}$ for the calendar day d (Equation (1)), where t_1, t_2 are the start and end date of the time series. Second, an equal probability is set for each element in $S_{x,d}\big|_{t_1}^{t_2}$ (Equation (2)) and the values are ordered from smallest to largest (Equation (3)). The cumulative probabilities $F_d(x)$ are calculated (Equation (4)) for each x value. Finally, the percentile values $q_p(F_d(x))$ are calculated (Equation (5)) for the cumulative probabilities from 0 to 1 every 0.05 steps, with q_0 as the minimum value, q_1 as the maximum. In most of the cases the q_i percentiles were $i \in [0.05, 0.95]$, linearly interpolated using the closest upper and lower values.

$$S_{x,d}\big|_{t_1}^{t_2} = \{x(t): t \in [t_1, t_2] \text{ and } \text{day}(t) = d\} \text{ or} \tag{1}$$

where $x = x(t)$: value of soil moisture at a time t, $S_{x,d}\big|_{t_1}^{t_2} \subset S_x\big|_{t_1}^{t_2}$, $t_1 = 1/2/1979$, and $t_2 = 12/31/2013$.

$$f_d(x) = P_d(X = x) = \frac{1}{n} \text{ and} \tag{2}$$

$$\text{Ordered}\ (S_{x,d}\big|_{t_1}^{t_2}) = O_{x,d}\big|_{t_1}^{t_2} \tag{3}$$

$$F_d(x) = P_d(X \le x_i) = \sum_{x(1)}^{x(i)} P_d(X = x_i) \qquad (4)$$

where $\{x: x(1), x(2), x(3), \ldots, x(n)\} = O_{x,d}\big|_{t_1}^{t_2}$.

$q_p =$

$$\begin{cases} x(1) \text{ if } p = 0 \\ \left(\dfrac{q_{p(U)} - q_{p(L)}}{F_d(x)_U - F_d(x)_L}\right)(F_d(x) - F_d(x)_L) + q_{p(L)} \text{ if } p \in [0.05, 0.95] \\ x(n) \text{ if } p = 1 \end{cases}$$

$$(5)$$

where $(q_{p(U)}, F_d(x)_U)$ is the closest upper value, $(q_{p(L)}, F_d(x)_L)$ is the closest lower value, and $q_{p(L)} < q_p < q_{p(U)}$.

Each grid cell in Texas has 365 CDFs, one for each calendar day d. The CDFs are the soil moisture values for the percentiles from 0 to 1 at 0.05 steps. The latest soil moisture results in NLDAS are compared with the CDFs values given the location and the day of the year. A percentile value is calculated where values close to zero are dryer than historic conditions and values close to one are wetter than historic conditions.

Figure 1 │ Components of the web application: client-side, server-side, and cloud development.

WEB APPLICATION

The web application architecture is structured with three underlying components (Figure 1): (1) client-side, (2) server-side, and (3) cloud. The client-side is the web application itself that displays the results of the statistical analysis through maps and charts. The server-side consists of two elements: a map layer that contains the geographic information and the NLDAS web services that deliver the most current data values. The cloud deployment component is where the application resides and where the temporal-spatial statistics are stored.

Figure 2 shows a web application prototype, which is accessible at http://texassoilmoisture.azurewebsites.net. The map can display (1) the latest soil moisture values in the NLDAS-Noah model, (2) the anomaly (i.e., difference from the mean), or (3) the percentile from the CDF for that particular day of the year. The top-right chart compares the latest value to the historic CDF, and the bottom-right chart show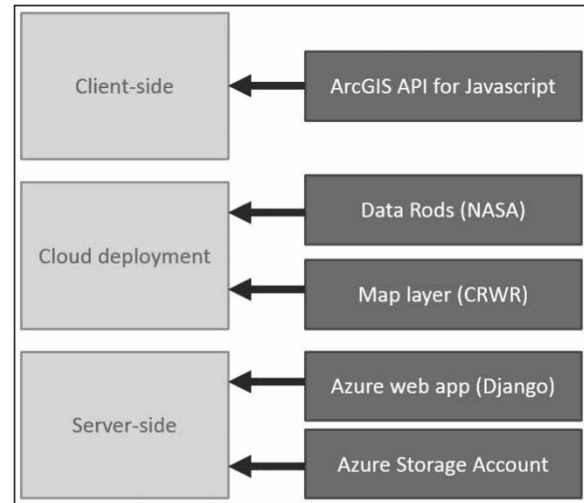s the variation in soil moisture over the previous 30 days. The charts displayed in Figure 2 for a grid cell north of Austin show that the soil moisture increased in the past days due to storm events, and had wetter conditions than historically normal for that day (5 February 2015).

Client-side

The client web application is deployed using a Python web framework known as Django (Django 2015). This web framework allows the deployment of complex applications that need access to databases, external servers and provide an HTML interface for the user. In this scheme, the user interacts with the web application by clicking on the map, resulting in click-events sending to Django the grid location identifier using the ArcGIS API for JavaScript (Esri 2015). This grid identifier is used as an input for two Python modules: one that retrieves data from the cloud storage platform, and another that collects the soil moisture time series for the previous 30 days from the NASA data rods web service. The results from these two requests are encoded in JSON and returned to the web client for geospatial visual representation and plotting statistical charts. Geospatial content is loaded in the client web application using the ArcGIS API for JavaScript, which provides a light and fast way of embedding map layers in web applications. Statistical plots are
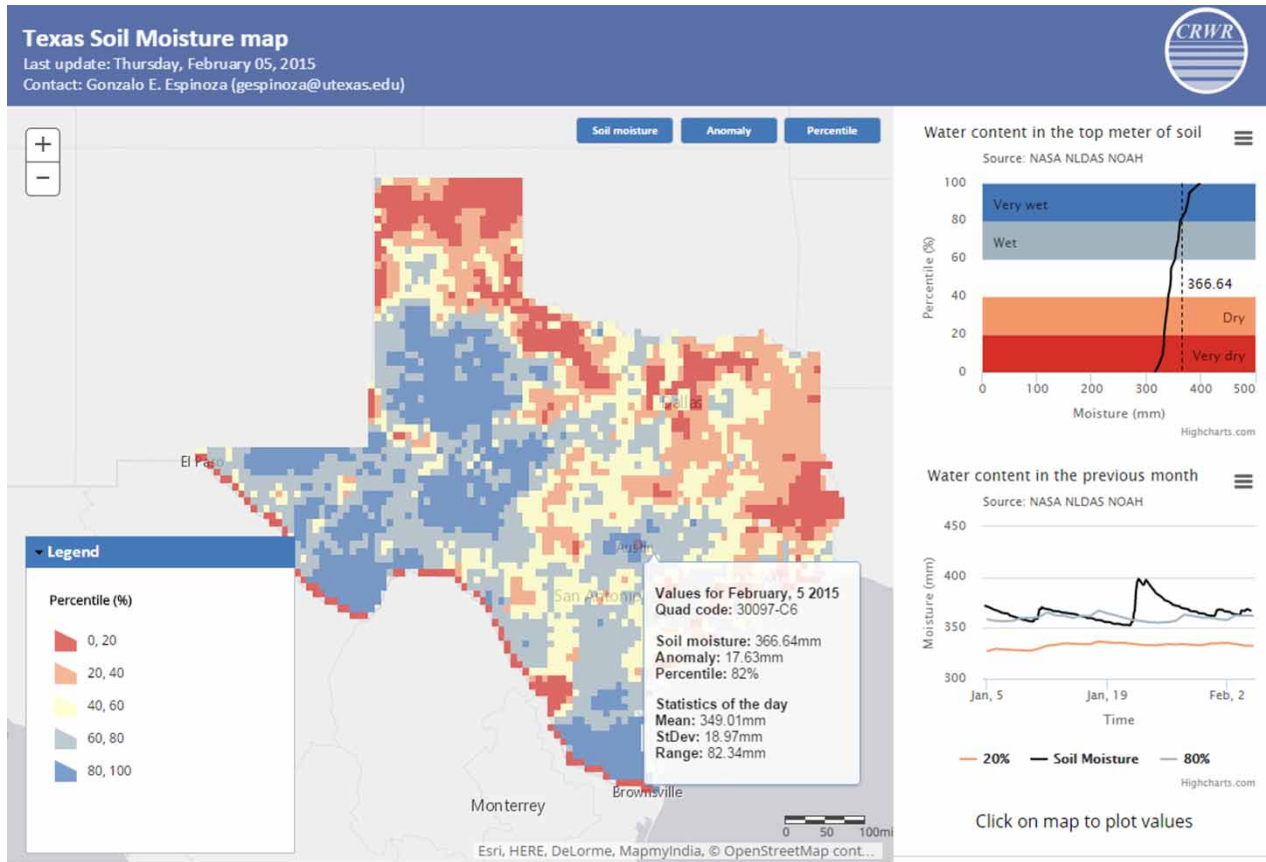
Figure 2 │ Web application displaying historic soil moisture CDFs and latest soil moisture conditions.

created using the HighCharts JavaScript library (Highcharts Developing Team 2015).

Server-side

Two servers support the web application: a geospatial server and a data server. The geospatial server is an instance of ArcGIS for Server running at CRWR, which provides the map layers used for the click-events interaction with the web application. The map layers on the geospatial server are web services exposed through a REST API that can be queried from the client-side to visualize them and provide user interaction. The map layers are used to identify the geographic locations, invoke the statistical analysis (through user interaction), and include the latest results for each cell. The maps also serve as an intermediate platform between the user and the cloud storage. The NASA data rods server provides the latest information and the time series data. It is accessed every time a user clicks on the map, and the real-time response allows the charts to be drawn instantaneously.

Cloud

The cloud component of the web application is deployed using Microsoft Azure (Microsoft 2015) and it includes two parts: web deployment and data storage. The web deployment component loads (1) the entire Django website, (2) the Python libraries, functions, and scripts, and (3) the JavaScript code. By doing this, the web application is available on the Internet through the prototype website (http://texassoilmoisture.azurewebsites.net/) and can benefit from the main features of cloud computing, such as supporting simultaneous multiple users or a fast performance between all the components of this web

application. The data cloud storage contains the results from the statistical analysis. The data are accessible through the Django application itself using the Python Azure SDK (Microsoft Azure 2015).

CASE STUDIES AND RESULTS

The CDFs computed are estimates of the spatial-temporal distribution of a land surface model (LSM) variable. Figure 3 displays the CDF of soil moisture and the actual values for 13 June 2014 at three locations (top to bottom): Austin, El Paso, and Houston. The empirical CDF (black line) shows
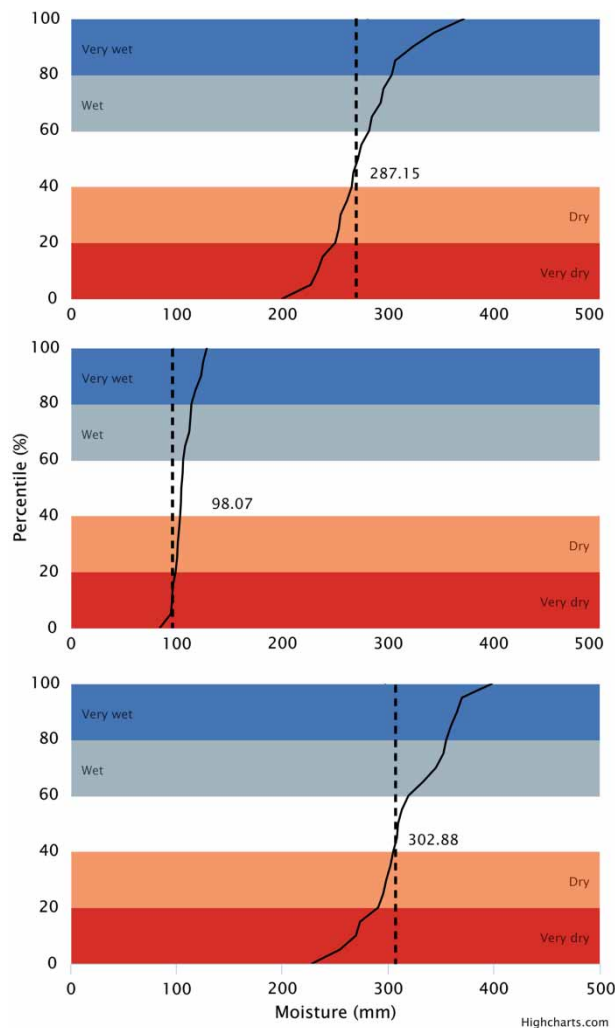
the spatial variation of soil moisture across Texas. The range of values varies in the three locations, by which Austin and Houston apparently have a larger value of soil moisture than El Paso. The shape of the line is sharper in El Paso and more extended in the two other cases. This is a measure of the variance of the values and how a storm event can impact the soil moisture conditions: a larger precipitation depth would be needed to come back to normal conditions from dry conditions in a more extended CDF. The latest conditions (dotted line) show how these values are compared to the historical ones for the same calendar day. Austin and Houston have normal conditions meanwhile El Paso is drier than usual. This methodology allows assessing and comparing soil moisture conditions relative to the ones in the past for the same location.

Figure 4 shows the variation of the percentiles across the year for the same three locations (from left to right: Austin, El Paso, and Houston). The seasonal variation across time is accurately represented. At El Paso, the range and magnitude of the values are considerably lower than in the other two cases and also show a different seasonality, having higher values during the summer. Meanwhile, at Austin and Houston, the summer season has the lower values of soil moisture and a larger range of values.

Figure 5 shows the CDF distributions for Austin at different times of the year. The plot shows how the distribution varies with time. The figure shows how the CDFs vary seasonally and that the percentile values are time dependent. The calculation of the CDFs for a given calendar day at each grid cell across Texas quantitatively describes the uncertainty of soil moisture values across time and space.

Case 1: Texas drought 2011

Figure 6 shows a soil moisture map of Texas for September 2011. The map illustrates the extremely dry conditions in Texas at that time. The vast majority of soil moisture values across Texas were below the 5th percentile.

Case 2: Halloween flood, Onion Creek 2013

Figure 7 shows the soil moisture conditions at the Onion Creek watershed located south of Austin just before the flood on 31 October 2013. The left plot shows the soil
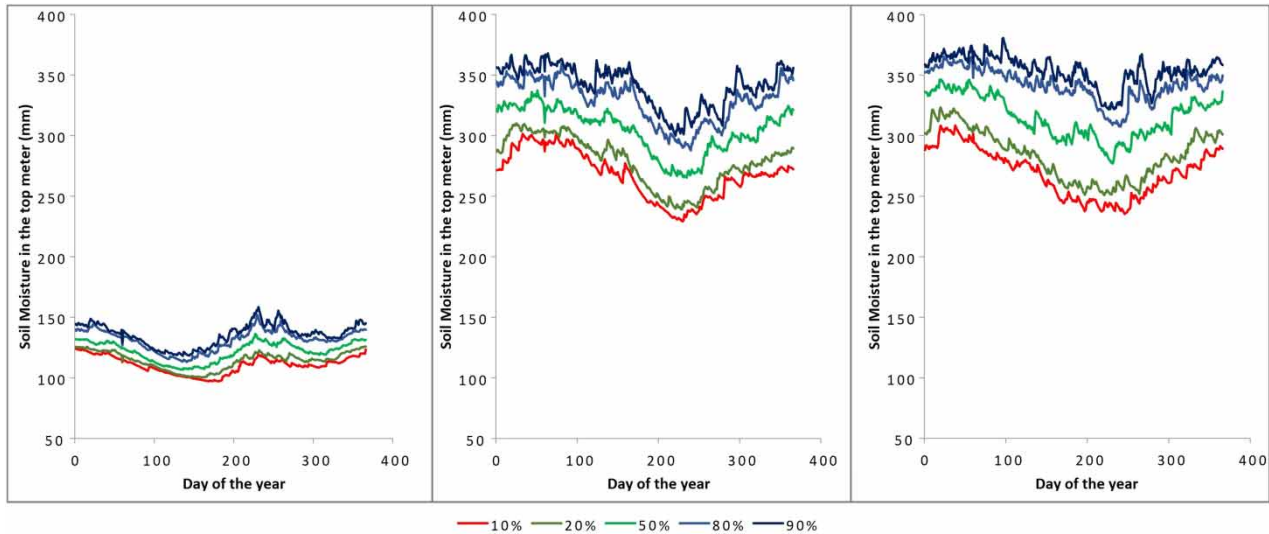

Figure 3 │ Soil moisture values and CDF distributions for 13 June 2014 at (top to bottom) Austin, El Paso, and Houston.

Figure 4 │ Variation of the CDFs across the year for the 10, 20, 50, 80, and 90 percentiles at (from left to right) Austin, El Paso, and Houston.
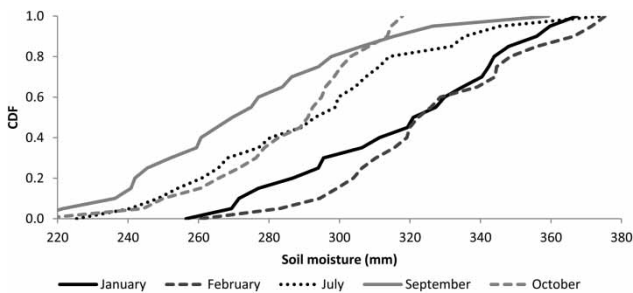

Figure 5 │ CDFs for different months (1st day of the month) at Austin.

moisture values, the 20th, and the 80th percentiles in the previous 30 days before the storm, identifying a significant increase in the soil moisture due to previous storms. The plot on the right shows the CDF and the pre-storm soil moisture that corresponds to an 87th percentile. These wet conditions indicate the potential for larger runoff volume in the event of another storm, which is exactly what happened.

DISCUSSION

The statistical analysis of soil moisture provides an insight of the pre-storm absolute water content in the soil. Nevertheless, more information is needed to properly describe an extreme event such as the 2013 Halloween flood in Austin. The estimation of percentage saturation from soil moisture would

be ideal, but additional soil information would be required. The Halloween flood was a complex extreme event in which different factors converged to increase its impact: the timing, magnitude, and distribution of the storm; the characteristics of the watershed; and human development and lack of flood control infrastructure. The statistical analysis of soil moisture presents a guide of when the soil is wetter than historic conditions and prone to produce more runoff if a storm occurs. However, soil moisture data alone are insufficient for a flood risk characterization. A study that expands this work to use more variables and linking the statistical analysis with real time data would be necessary.

For the Texas drought 2011 case, the soil moisture percentiles are a good parameter for understanding the persistent dry conditions. The daily time-step is small for this application and the same results can be achieved using a monthly time-step: reducing the data needed and the processing time. Moreover, similar statistical analyses for other NLDAS variables such as precipitation, runoff, and their interrelations are needed.

FURTHER WORK

The web application prototype is a seamless integration of data and geographic services. It is a cloud-based solution for displaying a large dataset in its geographic context. It
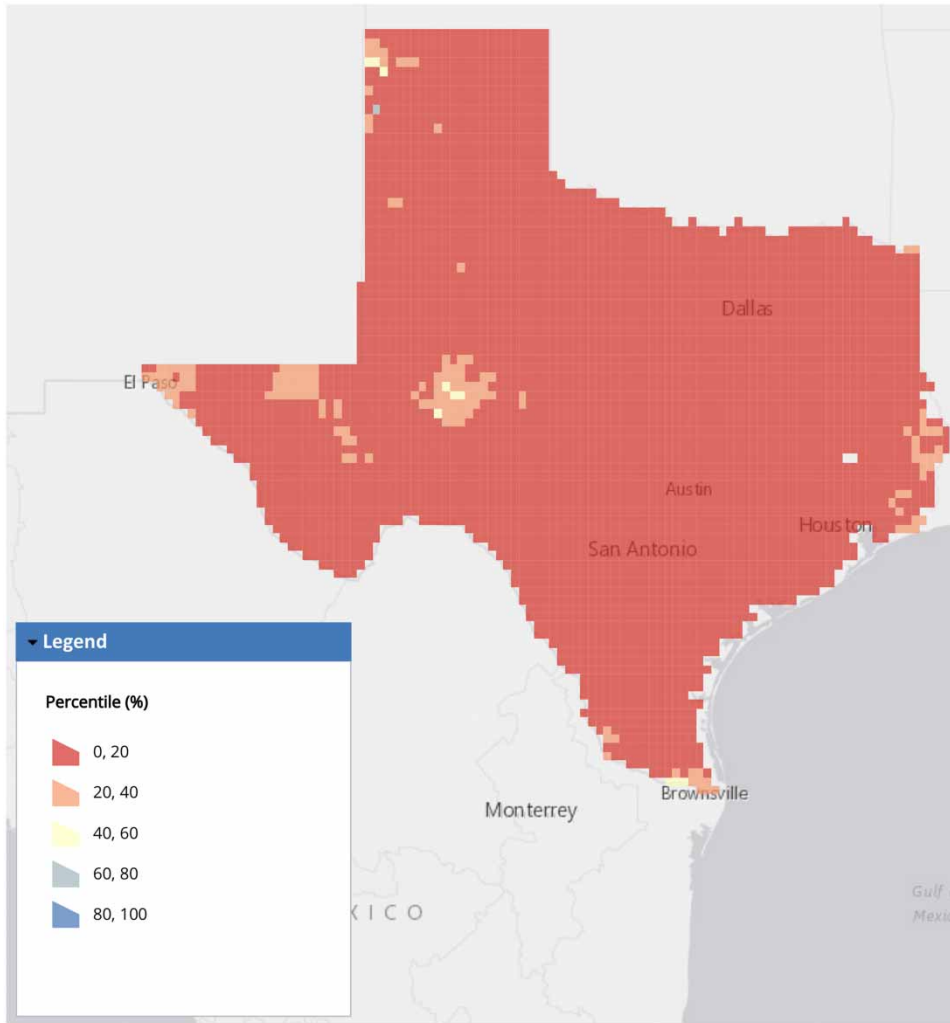
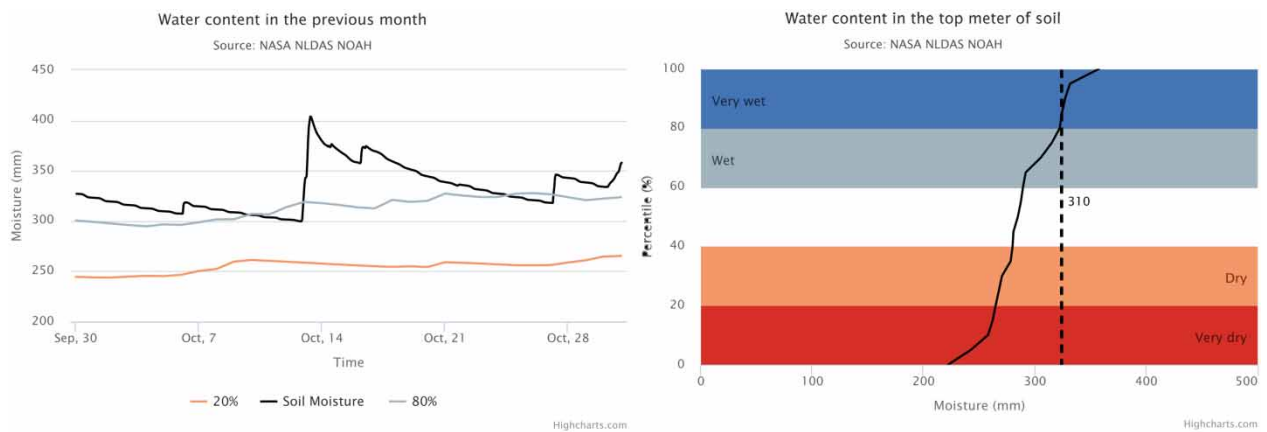**Figure 6** | Percentile distribution of soil moisture in Texas, September 2011.

Figure 7 | Pre-soil moisture conditions for the 'Halloween storm' at Onion Creek 31 October 2013.

shares its core ideas with Tethys (Jones *et al.* 2014) which is a Django-based platform to develop web applications, easing the web application development process. The following updates would be implemented using the Tethys platform:

- Improving performance of drawing grid (especially needed for national scale).
- Supporting selection of begin/end dates for time series (data rods explorer).
- Supporting display of other LSM outputs, such as precipitation rate, surface and subsurface runoff, and surface temperature.

ACKNOWLEDGEMENTS

REFERENCES

Beguería, S. 2005 Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *J. Hydrol.* **303** (1–4), 215–230.

Beniston, M., Stoffel, M., Harding, R., Kernan, M., Ludwig, R., Moors, E. & Tockner, K. 2012 Obstacles to data access for research related to climate and water: implications for science and EU policy-making. *Environ. Sci. Policy* **17**, 41–48.

Berman, F., Chien, A., Cooper, K., Dongarra, J., Foster, I., Gannon, D. & Wolski, R. 2001 The GrADS project: software support for high-level grid application development. *Int. J. High Perform. Comput. Appl.* **15** (4), 327–344.

Choi, J., Engel, B. A. & Farnsworth, R. L. 2005 Web-based GIS and spatial decision support system for watershed management. *J. Hydroinform.* **7** (3), 165–174.

Coles, S., Pericchi, L. R. & Sisson, S. 2003 A fully probabilistic approach to extreme rainfall modeling. *J. Hydrol.* **273** (1–4), 35–50.

Django. 2015 Django Version 1.5 Computer Software. Django Software Foundation. https://djangoproject.com.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V. & Tarpley, J. D. 2003 Implementation of Noah land surface model advances in the national centers for environmental prediction operational mesoscale Eta model. *J. Geophys. Res. Atmos.* **108** (D22), 8851. Doi:10.1029/2002JD003296.

Elshorbagy, A. & El-Baroudy, I. 2009 Investigating the capabilities of evolutionary data-driven techniques using the challenging estimation of soil moisture content. *J. Hydroinform.* **11** (3–4), 237. Doi:10.2166/hydro.2009.032.

El Adlouni, S., Bobée, B. & Ouarda, T. B. M. J. 2008 On the tails of extreme event distributions in hydrology. *J. Hydrol.* **355** (1–4), 16–33.

Esri. 2015 ArcGIS API for Javascript. https://developers.arcgis.com/javascript/jshelp/.

Funk, C., Michaelsen, J., Verdin, J., Artan, G., Husak, G., Senay, G. & Magadzire, T. 2003 The collaborative historical African rainfall model: description and evaluation. *Int. J. Climatol.* **23** (1), 47–66.

Gourbesville, P. 2009 Data and hydroinformatics: new possibilities and challenges. *J. Hydroinform.* **11** (3–4), 330. Doi:10.2166/hydro.2009.143.

Highcharts Developing Team 2015 Highcharts: Make your data come alive. http://www.highcharts.com/.

Husak, G. J., Michaelsen, J. & Funk, C. 2007 Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *Int. J. Climatol.* **27** (7), 935–944.

Jiang, P., Gautam, M. R., Zhu, J. & Yu, Z. 2013 How well do the GCMs/RCMs capture the multi-scale temporal variability of precipitation in the Southwestern United States? *J. Hydrol.* **479**, 75–85.

Jones, N. L., Nelson, E. J. & Manwaring, C. T. 2000 Managing temporal data in a comprehensive modeling environment. *J. Hydroinform.* **2** (2), 105–112.

Jones, N., Nelson, J., Swain, N., Christensen, S., Tarboton, D. & Dash, P. 2014 *Tethys: A Software Framework for Web-Based Modeling and Decision Support Applications.* In: International Environmental Modelling and Software Socierty (iEMSs). San Diego, CA.

Katz, R. W., Parlange, M. B. & Naveau, P. 2002 Statistics of extremes in hydrology. *Adv. Water Resour.* **25** (8–12), 1287–1304.

Kothyari, U. C., Ramsankaran, R., Sathish Kumar, D., Ghosh, S. K. & Mendiratta, N. 2010 Geospatial-based automated watershed modeling in Garhwal Himalaya. *J. Hydroinform.* **12** (4), 502.

Lakshmi, V. 2004 The role of satellite remote sensing in the Prediction of Ungauged Basins. *Hydrological Processes* **18** (5), 1029–1034. doi:10.1002/hyp.5520.

Lakshmi, V., Piechota, T., Narayan, U. & Tang, C. 2004 Soil moisture as an indicator of weather extremes. *Geophysical Research Letters* **31** (11) doi:10.1029/2004GL019930.

Liu, Y., Padmanabhan, A. & Wang, S. 2015 CyberGIS Gateway for enabling data-rich geospatial research and education. *Concurrency Comput. Pract. Exper.* **27** (2), 395–407.

Microsoft 2015 Microsoft Azure: Cloud Computing Platform & Services. https://azure.microsoft.com/en-us/.

Microsoft Azure 2015 Microsoft Azure SDK for Python. https://github.com/Azure/azure-sdk-for-python.

Mishra, S. 2009 Uncertainty and sensitivity analysis techniques for hydrologic modeling. *J. Hydroinform.* **11** (3–4), 282. Doi:10.2166/hydro.2009.048.

Mitchell, K. E. 2004 The multi-institution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.* **109** (D7), D07S90. Doi:10.1029/2003JD003823.

Rui, H., Strub, R., Teng, W., Vollmer, B., Mocko, D., Maidment, D. & Whiteaker, T. 2013 Enhancing Access to and Use of NASA Hydrological Data. AGU Fall Meeting (p. H11E-1200). http://disc.sci.gsfc.nasa.gov/additional/publications/year_2013/access_hydrological_data.pdf.

Xia, Y., Ek, M., Wei, H. & Meng, J. 2012 Comparative analysis of relationships between NLDAS-2 forcings and model outputs. *Hydrol. Process.* **26** (3), 467–474.